第27卷 第4期 工 程 数 学 学 报 Vol. 27 No. 4
2010年08月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Aug. 2010

文章编号:1005-3085(2010)04-0599-06

两两NQD序列下线性指数分布参数的经验Bayes检验*

王 亮， 师义民

(西北工业大学应用数学系，西安 710072)

摘 要: 本文讨论了线性指数分布参数的经验Bayes单边检验问题。对于两两NQD样本序列，利用核函数估计方法给出了检验函数。证明了该检验函数是渐近最优的，并在适当的条件下得到其收敛速度。

关键词: 两两NQD序列；核估计；经验Bayes检验；渐近最优性；收敛速度

分类号: AMS(2000) 62C12; 62F15 **中图分类号**: O212.1 **文献标识码**: A

1 引言

经验Bayes (Empirical Bayes, EB) 方法是Robbins于上世纪五六十年代提出的。对于EB估计和检验问题已有了相当多的研究。考虑到在实际问题中样本大都存在一定的相依关系及近年来线性指数分布在新产品，新工艺有效性检验等方面的重要意义，本文在两两NQD (Negative Quadrant Dependent, NQD)序列下，讨论线性指数模型参数的EB检验问题，利用概率密度核估计方法给出参数的EB检验函数，获得了该检验函数的渐近最优性及收敛速度。

首先给出两两NQD序列的定义。

定义1 称随机变量($r.v.$) X 和 Y 是NQD的，若对任意的 $x,\, y \in R$，有

$$P(X < x, Y < y) \le P(X < x)P(Y < y),$$

称 $r.v.$ 序列 $\{X_n : n = 1, 2, \cdots\}$ 是两两NQD的，若对任意的 $i \neq j$ 有 X_i 与 X_j 是NQD的。

两两NQD序列是由Lemman[1] 提出的，可以看出它一类应用十分广泛的相依序列，在可靠性理论，多元分析等领域中都十分常见，许多负相依序列都是在它基础上生成的，如著名的负相依(NA)序列就是它的特殊情况之一，因此对于两两NQD序列的讨论就显得更加重要。在这方面已有了一些研究，如Mutula[2] 获得了同分布的两两NQD序列的Kolmogorov型强大数定律。

考虑线性指数模型，设随机变量 X 的条件密度函数为

$$f(x\,|\,\theta) = (\mu x + \theta)\exp[-\theta x - \mu x^2/2], \tag{1}$$

其中 $\theta,\, \mu$ 为参数且 μ 已知，样本空间 $\Omega = \{x : x > 0\}$，参数空间

$$\Theta = \Big\{\theta > 0 : \int_\Omega f(x\,|\,\theta)\mathrm{d}x = 1\Big\}.$$

收稿日期: 2008-11-04. **作者简介**: 王亮(1983年2月生)，男，博士. 研究方向：应用概率统计与可靠性理论.
***基金项目**: 国家自然科学基金(70471057)；陕西省教育厅自然科学基金(03JK065).

假设参数θ的先验分布为$G(\theta)$，则X的边缘密度及其导数为

$$f(x)=\int_\Theta f(x\,|\,\theta)\mathrm{d}G(\theta),\quad f'(x)=\int_\Theta(\mu-\mu^2x^2-2\mu\theta x-\theta^2)\mathrm{e}^{-\theta x-\mu x^2/2}\mathrm{d}G(\theta).$$

考虑单边检验问题，$H_0:\theta\le\theta_0\leftrightarrow H_1:\theta>\theta_0$，其中$\theta_0$是给定常数。取线性损失函数

$$L(\theta,d_0)=\max\big\{c(\theta-\theta_0),0\big\},\quad L(\theta,d_1)=\max\big\{c(\theta_0-\theta),0\big\},$$

这里c为正常数，$D=\{d_0,d_1\}$是行动空间，d_i表示接受H_i，$i=0,1$。

取随机化判决函数$\delta(x)=P\{$接受$H_0\,|X=x\}$，则$\delta(x)$的风险函数为

$$\begin{aligned}R\big(\delta(x),G\big)&=\int_\Theta\int_\Omega\big[L(\theta,d_0)\delta(x)+L(\theta,d_1)(1-\delta(x))\big]f(x\,|\,\theta)\mathrm{d}x\mathrm{d}G(\theta)\\&=\int_\Omega\alpha(x)\delta(x)\mathrm{d}x+C_G,\end{aligned}$$

这里

$$C_G=\int_\Theta L(\theta,d_1)\mathrm{d}G(\theta),$$

$$\alpha(x)=\int_\Theta c(\theta-\theta_0)f(x\,|\,\theta)\mathrm{d}G(\theta)=u(x)f(x)+v(x)f'(x)+w(x)p_G(x),$$

其中

$$u(x)=-c(\mu x+\theta_0),\quad v(x)=-c,\quad w(x)=c\mu,\quad p_G(x)=\int_\Theta\mathrm{e}^{-\theta x-\mu x^2/2}\mathrm{d}G(\theta).$$

进而θ的EB检验函数为$\delta_G=1\cdot I_{(\alpha(x)\le0)}$，相应的Bayes风险为

$$R(\delta_G,G)=\inf_{x\in R}R(\delta(x),G)=\int_\Omega\alpha(x)\delta_G(x)\mathrm{d}x+C_G.\tag{2}$$

当$G(\theta)$已知且$\delta(x)=\delta_G(x)$时，Bayes风险$R(\delta_G,G)$可以取到，但通常$G(\theta)$未知，所以检验函数$\delta_G(x)$不可用，从而引入EB方法。

2 两两NQD序列经验Bayes检验函数的构造

设$(X_1,\theta_1),\cdots,(X_n,\theta_n)$为同分布两两NQD序列。本文假定$f(x)\in C_{s,a}$，这里$C_{s,a}$表示$R$中的一类$s$阶导数存在、连续且绝对值不超过$a$的概率密度函数族，$s\ge3$为正整数。

以下逐步给出检验函数$\alpha(x)$估计量的构造。

设$K_i(x)\,(i=0,1)$是区间$(0,1)$上有界可测函数，几乎处处可微，在$(0,1)$之外为零且满足

$$(L)\quad\frac{1}{t!}\int_0^1u^tK_i(u)\mathrm{d}u=\begin{cases}1, & t=i,\\0, & t\ne i,\quad t=0,1.\end{cases}$$

下面给出概率密度函数$f(x)$及其导数$f'(x)$的核函数估计为

$$f_n(x)=\frac{1}{nh_n}\sum_{i=1}^nK_0\Big(\frac{x-X_i}{h_n}\Big),\quad f_n'(x)=\frac{1}{nh_n^2}\sum_{i=1}^nK_1\Big(\frac{x-X_i}{h_n}\Big),$$

其中 $\{h_n\}$ 为正数序列，且 $h_n \to 0\,(n \to \infty)$。

同时，由于

$$p_G(x) = \int_x^{+\infty} f(x)\mathrm{d}x = E\left[I_{(X>x)}\right],$$

故 $p_G(x)$ 的估计为

$$\bar{p}_G(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(X_i>x)},$$

从而检验函数 $\alpha(x)$ 的估计量 $\alpha_n(x)$ 为 $\alpha_n(x) = u(x)f_n(x) + v(x)f_n'(x) + w(x)\bar{p}_G(x)$，相应的EB检验函数定义为 $\delta_n(x) = 1 \cdot I_{(\alpha_n(x)\leq 0)}$。令 E_n 表示对 $X_1, X_2, \cdots, X_n$ 的联合分布取期望，则 $\delta_n(x)$ 的全面Bayes风险为

$$R(\delta_n, G) = E_n\Big[\int_\Omega \alpha(x)\delta_n(x)\mathrm{d}x + C_G\Big] = \int_\Omega \alpha(x)E_n\left[\delta_n(x)\right]\mathrm{d}x + C_G. \tag{3}$$

定义2 设 $\mathscr{F}$ 表示参数 θ 的先验分布族，若对任意的 $G(\theta) \in \mathscr{F}$，有

$$\lim_{n\to\infty} R(\delta_n, G) = R(\delta_G, G),$$

则称EB检验函数 $\{\delta_n(x)\}$ 为渐近最优的；若 $R(\delta_n, G) - R(\delta_G, G) = O(n^{-p})\,(p > 0)$，则称EB检验函数具有收敛速度 $O(n^{-p})$。

3 若干引理

本节给出若干引理。这里令 $M, M_1, \cdots$ 表示不同常数，即使同一表达式中也可取不同值。

引理1[1] 设随机变量 X 和 Y 是NQD的，若 r, s 同时为非降(非增)实函数，则 $r(X)$, $s(Y)$ 仍为NQD的。

由定义知 $K_i(x)\,(i = 0,1)$ 是实数域 R 上的有界变差函数，根据实变函数理论，存在 R 上的单调有界变差函数 $K_i^1(x)$, $K_i^2(x)$, $i = 0,1$，使得 $K_i(x) = K_i^1(x) - K_i^2(x)$, $i = 0,1$。

因为 $\{X_i : i = 1,2,\cdots,n\}$ 为两两NQD序列，由引理1知

$$\Big\{\frac{x - X_i}{h_n}\Big\}_{i=1}^{n}$$

为两两NQD序列。进一步有

$$\Big\{\frac{1}{h_n}K_m^i\Big(\frac{x - X_j}{h_n}\Big)\Big\}_{j=1}^{n} \triangleq \big\{Q_{mj}^i\big\}_{j=1}^{n}, \quad i = 1,2, \quad m = 0,1; \quad \big\{I_{(X_i>x)}\big\}_{i=1}^{n} \triangleq \{Q_i^3\}_{i=1}^{n}$$

为同分布两两NQD序列，从而

$$f_n(x) = \frac{1}{n}\sum_{j=1}^{n} Q_{0j}^1 - \frac{1}{n}\sum_{j=1}^{n} Q_{0j}^2, \quad f_n'(x) = \frac{1}{n}\sum_{j=1}^{n} Q_{1j}^1 - \frac{1}{n}\sum_{j=1}^{n} Q_{1j}^2, \quad \bar{p}_G(x) = \frac{1}{n}\sum_{j=1}^{n} Q_j^3,$$

即 $\alpha_n(x)$ 的各分量可以分解为两两NQD序列和的形式。

引理2[3] 设 $\{X_i : i = 1,2,\cdots,n\}$ 为两两NQD序列

$$EX_i^2 < +\infty, \quad T_j(k) = \sum_{i=j+1}^{j+k} (X_i - EX_i), \quad j \geq 1,$$

则

$$E\big(T_j(k)\big)^2 \le \sum_{i=j+1}^{j+k} EX_i^2.$$

引理3[4] 令 $R(\delta_G,G)$, $R(\delta_n,G)$ 如前式 (2), (3) 定义，则

$$0 \le R(\delta_n,G) - R(\delta_G,G) \le c\int_\Omega |\alpha(x)|P\big(|\alpha_n(x)-\alpha(x)| \ge |\alpha(x)|\big)\mathrm{d}x.$$

引理4 设 $\{X_i : i=1,2,\cdots,n\}$ 为两两NQD序列，核函数条件 (L) 满足，则

1) 当 $n\to 0$, $h_n\to 0$, $nh_n^2\to+\infty$ 时，$E_n|f_n(x)-f(x)|^2\to 0$;

2) 当 $n\to 0$, $h_n\to 0$, $nh_n^4\to+\infty$ 时，$E_n|f_n'(x)-f'(x)|^2\to 0$;

3) 当 $n\to 0$ 时，$E_n|\bar{p}_G(x)-p_G(x)|^2\to 0$。

证明 1) 由 C_r 不等式

$$E_n|f_n(x)-f(x)|^2 \le M_1E_n\big[f_n(x)-E_nf_n(x)\big]^2 + M_2\big[E_nf_n(x)-f(x)\big]^2.$$

一方面

$$\big|E_nf_n(x)-f(x)\big| = \Big|\int_\Omega K_0(t)\big[f(x-th_n)-f(x)\big]\mathrm{d}t\Big|.$$

因为 $f(x)\in C_{s,a}$，由Taylor展式有

$$\begin{aligned} f(x-th_n) &= f(x)+\frac{f'(x)}{1!}(-th_n)+\cdots \\ &\quad +\frac{f^{(s-1)}(x)}{(s-1)!}(-th_n)^{s-1}+\frac{f^s(\xi)}{s!}(-th_n)^s,\quad \xi\in(x-th_n,x), \end{aligned}$$

利用核函数性质可知

$$\big|E_nf_n(x)-f(x)\big| \le M\Big|\int_\Omega K_0\frac{f^s(\xi)}{s!}(-th_n)^s\mathrm{d}t\Big| \le Mh_n^s.$$

另一方面，由引理2可知

$$\begin{aligned} E_n\big[f_n(x)-E_nf_n(x)\big]^2 &\le 2\sum_{m=1}^{2}E_n\Big\{\frac{1}{n}\sum_{j=1}^{n}\big(Q_{0j}^m-EQ_{0j}^m\big)\Big\}^2 \\ &\le \frac{2}{n^2}\sum_{m=1}^{2}\sum_{j=1}^{n}E\big\{(Q_{0j}^m-EQ_{0j}^m)\big\}^2 \le \frac{2}{n}E\big[Q_{01}^1\big]^2+\frac{2}{n}E\big[Q_{01}^2\big]^2 \le \frac{M}{nh_n^2}, \end{aligned}$$

因此结论1)成立且

$$\big|E_nf_n(x)-f(x)\big| \le Mh_n^s,\quad E_n\big|f_n(x)-E_nf_n(x)\big|^2 \le \frac{M}{nh_n^2}.$$

2) 类似证明1)，结论2)成立且

$$\big|E_nf_n'(x)-f'(x)\big| \le Mh_n^{s-1},\quad E_n\big|f_n'(x)-E_nf_n'(x)\big|^2 \le \frac{M}{nh_n^4}.$$

3) 类似证明1)，结论3)成立且

$$\big|E_n\bar{p}_G(x)-\bar{p}_G(x)\big| = 0,\quad E_n\big|\bar{p}_G(x)-E_n\bar{p}_G(x)\big|^2 \le \frac{M}{n}.$$

4　经验Bayes检验函数的渐近最优性及收敛速度

本节给出文章主要结果，即EB检验函数的渐近最优性及其收敛速度。

定理1 (渐近最优性)　设$\{X_i : i = 1, 2, \cdots, n\}$为两两NQD序列，核函数条件$(L)$满足，且

$$\int_{\Theta} |\theta| \mathrm{d}G(\theta) < +\infty,$$

则当$n \to \infty$, $h_n \to 0$, $nh_n^4 \to \infty$时，有

$$\lim_{n\to\infty} R(\delta_n, G) = R(\delta_G, G).$$

证明　令

$$D_n(x) = |\alpha(x)| P\big(|\alpha_n(x) - \alpha(x)| \geq |\alpha(x)|\big),$$

则$D_n(x) \leq |\alpha(x)|$。由$\alpha(x)$定义，有

$$\int_{\Omega} |\alpha(x)| \mathrm{d}x \leq c \int_{\Theta} |\theta| \mathrm{d}G(\theta) + c\theta_0 < +\infty,$$

因此由控制收敛定理及引理3得

$$\lim_{n\to\infty} R(\delta_n, G) - R(\delta_G, G) \leq c \int_{\Omega} \lim_{n\to\infty} D_n(x) \mathrm{d}x.$$

由Markov不等式及$\alpha(x)$定义有

$$D_n(x) \leq |u(x)| E_n |f_n(x) - f(x)| + |v(x)| E_n |f_n'(x) - f'(x)| + |w(x)| E_n |\bar{p}_G(x) - p_G(x)|,$$

根据定理条件及引理4得，对任意的$x \in \Omega$，有$\lim\limits_{n\to\infty} D_n(x) = 0$，即

$$\lim_{n\to\infty} R(\delta_n, G) = R(\delta_G, G).$$

定理2 (收敛速度)　设$\{X_i : i = 1, 2, \cdots, n\}$为两两NQD序列，核函数条件$(L)$满足，且对任意的$\lambda \in (0, 1)$

$$\int_{\Omega} |\alpha(x)|^{1-\lambda} |u(x)|^{\lambda} \mathrm{d}x < \infty, \quad \int_{\Omega} |\alpha(x)|^{1-\lambda} \mathrm{d}x < \infty,$$

则当$h_n = n^{-\frac{1}{2s+3}}$时，有

$$R(\delta_n, G) - R(\delta_G, G) = O\Big(n^{-\frac{\lambda(2s-5)}{2(2s+3)}}\Big).$$

证明　由Markov不等式及引理3可知

$$\begin{aligned} R(\delta_n, G) - R(\delta_G, G) \leq\ & M_1 E_n \big|f_n(x) - f(x)\big|^{\lambda} + M_2 E_n \big|f_n'(x) - f'(x)\big|^{\lambda} \\ & + M_3 E_n \big|\bar{p}_G(x) - p_G(x)\big|^{\lambda}, \end{aligned}$$

由条件$0 < \lambda < 1 < 2$及引理4可得

$$\begin{aligned} E_n \big|f_n(x) - f(x)\big|^{\lambda} \leq\ & \big[E_n |f_n(x) - f(x)|^2\big]^{\lambda/2} \\ \leq\ & \big[M_1 |E_n f_n(x) - f(x)|^2 + M_2 E_n |f_n(x) - E_n f_n(x)|\big]^{\lambda/2} \\ \leq\ & M \Big[\frac{1 + nh_n^{2s+2}}{nh_n^2}\Big]^{\lambda/2}, \end{aligned}$$

类似可得

$$E_n\left|f_n'(x)-f'(x)\right|^\lambda \le M\left[\frac{1+nh_n^{2s+2}}{nh_n^4}\right]^{\lambda/2}, \quad E_n\left|\bar{p}_G(x)-\bar{p}_G(x)\right|^\lambda \le M\left[\frac{1}{n}\right]^{\lambda/2}.$$

综上，当 $h_n=n^{-\frac{1}{2s+3}}$ 时，有

$$\begin{aligned} R(\delta_n,G)-R(\delta_G,G) &\le M\left[\frac{1+nh_n^{2s+2}}{nh_n^2}\right]^{\lambda/2}+M\left[\frac{1+nh_n^{2s+2}}{nh_n^4}\right]^{\lambda/2}+M\left[\frac{1}{n}\right]^{\lambda/2} \\ &\le Mn^{-\frac{\lambda(2s-5)}{2(2s+3)}}, \end{aligned}$$

即

$$R(\delta_n,G)-R(\delta_G,G)=O\left(n^{-\frac{\lambda(2s-5)}{2(2s+3)}}\right).$$

参考文献：

[1] Lehmann E L. Some concepts of dependence[J]. Ann Math Statist, 1966, 37(5): 1137-1153

[2] Matula P. A note on the almost sure convergence of sums of negatively dependent random variables[J]. Statist Probab Lett, 1992, 15: 209-213

[3] 吴群英. 两两NQD序列的收敛性质[J]. 数学学报，2002, 45(3): 617-624
Wu Q Y. Convergence properties of pairwise NQD random sequences[J]. Acta Mathematica Sinca, 2002, 45(3): 617 - 624

[4] John M V Jr, Van Ryzan J. Convegence rates in empirical Bayes two-action problems II: continuous case[J]. Ann Math Statist, 1972, 43: 934-937

[5] 王立春. 刻度参数的贝叶斯经验贝叶斯估计[J]. 工程数学学报，2007, 24(5): 795-800
Wang L C. Bayes empirical Bayes estimation for scale parameter[J]. Chinese Journal of Engineering Mathemmatics, 2007, 24(5): 795 - 800

[6] Prakasa Rao B L S. Nonparametric Functional Estimation[M]. New York: Academic Press, 1983

[7] Wei L S. The consistency for the kernel-type density estimation in the case of NA samples[J]. J Sys Sci & Math Scis, 2001, 21(1): 79-87

[8] 凌能祥，许昌满，彭小智. NQD样本下密度函数核估计的相合性[J]. 合肥工业大学学报(自然科学版)，2008, 31(2): 287-290
Ling N X, Xu C M, Peng X Z. Consistency of the kernel estimator of the density function for pairwise negative quadrant dependent sequences[J]. Journal of Hefei University of Technology (Natural Science), 2008, 31(2): 287-290

Empirical Bayes Test for Parameters of the Linear Exponential Distribution with NQD Random Samples

WANG Liang, SHI Yi-min

(Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072)

Abstract: In this paper, the one-sided test problem for parameters of the linear exponential families was investigated through the empirical Bayes (EB) approach. The EB test function was constructed by using the kernel-type estimation for Negative Quadrant Dependent (NQD) samples, and the convergence rates for the EB test function was obtained under suitable conditions. The results show that the EB test function is asymptotically optimal.

Keywords: NQD samples; kernel estimation; empirical Bayes test; asymptotic optimality; convergence rates

Received: 04 Nov 2008. **Accepted:** 24 Dec 2009.

Foundation item: The National Natural Science Foundation of China (70471057); the Natural Science Foundation of the Education Department of Shaanxi Province (03JK065).